1. (currently amended) A method for assembly of a plurality of reads from a genomic region, the method comprising the steps of:

(a) providing a plurality of reads from a genomic region;

(b) for each of said plurality of reads, indexing a plurality of read subsequences according to read number;~~for each of the plurality of reads;, each subsequence having an associated read with which it corresponds;~~

(c) extracting from the indexed subsequences a plurality of read pairs that have a ~~predetermined~~selected number of subsequences in common; and

(d) merging the read pairs along a continuum.

2. (currently amended) The method of claim 1, wherein said step (a) comprises ~~comprising the step of providing a plurality of reads~~ providing a plurality of reads generated from sequencing both ends of a plurality of DNA segments, each read ~~having~~ being associated with linking information comprising an associated orientation relative to a read from an opposite end of the DNA segment, and an associated distance from the read on the opposite end of the DNA segment.

3. (currently amended) The method of claim 1, ~~comprising~~ wherein~~the step of providing a plurality reads~~ said step (a) comprises providing a plurality of reads that are reverse complements of a plurality of reads provided by sequencing both ends of a plurality of DNA segments.

4. (currently amended) The method of claim 1, further comprising, after said step (b), ~~comprising the step of~~ sorting the indexed read subsequences alphabetically by subsequence.

5. (currently amended) The method of claim 1, further comprising~~the step of~~ discarding read subsequences having more than a cutoff number of occurrences from the plurality of indexed subsequences.

6. (currently amended) The method of claim 1, wherein each of said plurality of read subsequences has the same selected length.~~comprising the step of indexing a plurality of read subsequences of a predetermined length for each of the plurality of reads.~~

7. (currently amended) The method of claim 6, wherein the ~~predetermined~~selected length for each of the plurality of reads is between about 12 and about 32 bases long.

8. (currently amended) The method of claim 1, wherein said step (b) further comprises indexing the indexed subsequences by ~~comprising the step of indexing a plurality of subsequences for each read, the index comprising for each subsequence an associated read and~~ an associated starting position on the read ~~with which it corresponds~~.

9. (currently amended) The method of claim 1, wherein said step (d) comprises ~~comprising the step of performing alignments on~~aligning, according to sequence similarity, the plurality of read pairs having a ~~predetermined~~selected number of subsequences in common.

10. (currently amended) The method of claim 9~~8~~, wherein the step of aligning further comprises ~~comprising the steps of:~~
~~performing alignments on the plurality of read pairs having a predetermined number of subsequences in common; and~~
~~using~~ comparing the associated position on the reads with which the subsequences correspond to verify overlap.

11. (currently amended) The method of claim 2, further comprising: ~~the step of using the linking information associated with the reads to~~
determining linking information for said plurality of reads, said linking information comprising relative orientation of the reads and approximate distances between reads; and

determining linking information for said merged read pairs, said linking information comprising orientation of merged read pairs and approximate distances between merged read pairs; and

comparing the plurality of reads and merged reads for consistency~~confirm that the merged pairs are merged correctly~~.

12. (currently amended) The method of claim 2, further comprising: ~~the step of using the associated linking information to~~

determining linking information for said plurality of reads, said linking information comprising relative orientation of the reads and approximate distances between reads; and

determining linking information for said merged read pairs, said linking information comprising orientation of merged read pairs and approximate distances between merged read pairs; and

identifying an ambiguity in the merged reads by comparing the linking information of the merged read pairs with the linking information of said plurality of reads.

13. (currently amended) The method of claim 12, further comprising the step of identifying a repeat region and a set of unique regions.

14. (currently amended) The method of claim 13, further comprising the step of linking pairs of unique regions ~~using~~according to the linking information associated with the reads in the unique regions.

15. (currently amended) The method of claim 14, further comprising the step of inserting the repeat region between each linked pair of unique regions with which the repeat region corresponds.

16. (currently amended) The method of claim 13, further comprising the step of merging linked pairs of unique regions ~~using~~according to the linking information associated with the reads in the unique regions.

17. (currently amended) A method for assembly of merged reads from a genomic region, the method comprising the steps of:

providing one or more sets of merged reads from a genomic region comprising a set of reads ~~having~~ being associated with linking information;

~~using~~ comparing the linking information of said plurality of reads with the linking information of said merged reads ~~the associated linking information~~ to identify an ambiguity in the merged reads;

identifying a repeat region and a set of unique regions; and

linking pairs of unique regions ~~using~~ according to the linking information associated with the reads in the unique regions.

18. (previously presented) The method of claim 17, comprising the step of inserting the repeat region between each linked pair of unique regions with which the repeat region corresponds.

19. (currently amended) The method of claim 17, comprising the step of merging linked pairs of unique regions ~~using~~ according to the linking information associated with the reads in the unique regions.

20. (currently amended) An article of manufacture ~~having~~ comprising computer-readable program means embodied thereon for assembly of a plurality of reads from a genomic region, the article comprising:

computer-readable program means for providing a plurality of reads from a genomic region;

computer-readable program means for indexing, by an associated starting position on the read, a plurality of read subsequences for each of the plurality of reads, each subsequence having an associated read with which it corresponds;

computer-readable program means for extracting from the indexed subsequences a plurality of read pairs that have a ~~predetermined~~ selected number of subsequences in common; and

computer-readable program means for merging the read pairs along a continuum.

21. (currently amended) An article of manufacture ~~having~~ comprising computer-readable program means embodied thereon for assembly of merged reads from a genomic region, the article comprising:

computer-readable program means for providing one or more sets of merged reads from a genomic region comprising a set of reads ~~having~~ with associated linking information;

computer-readable program means for comparing the linking information of said plurality of reads with the linking information of said merged reads ~~using the associated linking information~~ to identify one or more ambiguities in the merged reads;

computer-readable program means for identifying a repeat region and a set of unique regions; and

computer-readable program means for linking pairs of unique regions ~~using~~ according to the linking information associated with the reads in the unique regions.

In the Drawings

The attached sheet of drawings includes changes to Figure 3. This sheet, which includes Figures 3A and 3B, replaces the original sheet including Figure 3. The two tables previously labeled jointly as Figure 3 are relabeled as Figures 3A and 3B as the Examiner requested. Applicants respectfully request the Examiner now enter the previously filed Figures 4, 7, and 8, filed with the May 31,2002 Preliminary Amendment, as the Examiner noted they were acceptable. No new matter is added.

**Attachments:** Replacement Sheets

Annotated Sheets Showing Changes

## Basis for the Amendments

Specification

The specification was amended, as described in the previous section, to make it consistent with the legend used by the drawings. Applicants also corrected a clerical error paragraph 0051 so that the three-letter example sequence of CGC in the third to last sentence is now GCG. The new sequence was referenced according to start position on the two example reads and, following the examples, one can determine that the sequence of GCG is the correct sequence. Lastly, Applicants corrected a typographical error in paragraph 0043. No new matter is added by these amendments.

Claims

Applicants amended independent claims 1 and 20 to recite that the "index for the plurality of read subsequences is according to read number". This change is supported in the drawings in Figure 2, "Table of Subsequences".

Applicants amended claim 6 to clarify that the plurality of read subsequences be of a uniform, selected length, not that they are indexed according to length. Figure 2, "Table of Subsequences", displays a listing of a plurality of subsequences of identical length. In paragraph 0040 of the specification, the first sentence reads "For the purposes of illustration, the subsequences in Figure 2 are 4 nucleotides long". Further support can be found in example given in the specification at paragraph 0071: "An index of 700,000,000 subsequences 24 bases long was generated..."

Claim 8 is amended to recite that the subsequences are further indexed by starting position on the read. Support for the amendment is found in Figure 3A.

Claim 4 is amended to specify that sorting criteria for the subsequences is alphabetical by subsequence. Support for this amendment can be found in Figure 3A.

Claims 2, 17, and 21 are amended to more clearly recite that each read is "associated with linking information". Claims 2 and 17 are also amended to recite that reads are associated with linking and distance information. This change is supported in paragraphs 0007 and 0032 of the specification.